4SL4 Assignment 4

Aaron Pinto
400190637
pintoa9

The plot for all 6 training methods is shown below.

The decision tree classifier was the worst overall with the highest training error.

The Bagging classifier had a basically flat error curve, around 0.061.

The Random Forest classifier was the best performer overall and the best performer at nearly all of the number of predictors. It started off weak, and then improved until around the number of predictors reached 350, and then got worse by a very slight amount, until it flatlined past 1150 predictors.

The Adaboost classifier with decision stumps (#1) showed improvement until 350 predictors, and then it had a substantial increase in error starting at 700 predictors, before it plateaud around 0.066, making it the 2nd worst classifier at the max number of predictors.

The Adaboost classifier with decision trees with at most 10 leaves (#2) had a large variance in test error for the beginning 500 number of predictors before dropping to around 0.053 at 700 predictors. It then spiked at 900 predictors before dropping generally (with a couple more spikes) to around 0.050 where it remained till the max number of predictors.

The Adaboost classifer with no restriction on the depth or node number (#3) was also essentially a flat line with very slight increases in error, and the 3rd best performer at the end.

As the number of predictors becomes larger, the Random Forest classifier and the Adaboost #2 classifier end up essentially tie-ing. The Random Forest classifier is the best predictor except for between 1600-1900 predictors and 2100-2300 predictors, where it's beaten by the Adaboost #2 classifier.

The ensemble methods are somewhat close in performance, being separated by 2% accuracy best-to-worst. The decision tree classifier is the outlier out of all the classifiers here at around 91.8% accuracy. The best performance at any number of predictors is the Random Forest classifier at around 95.2% accuracy at 350 predictors. Out of the ensemble methods, the worst performance is the Adaboost #1 classifier at around 93.2% accuracy at 1400, 1900, 1950, 2350, and 2400 predictors.
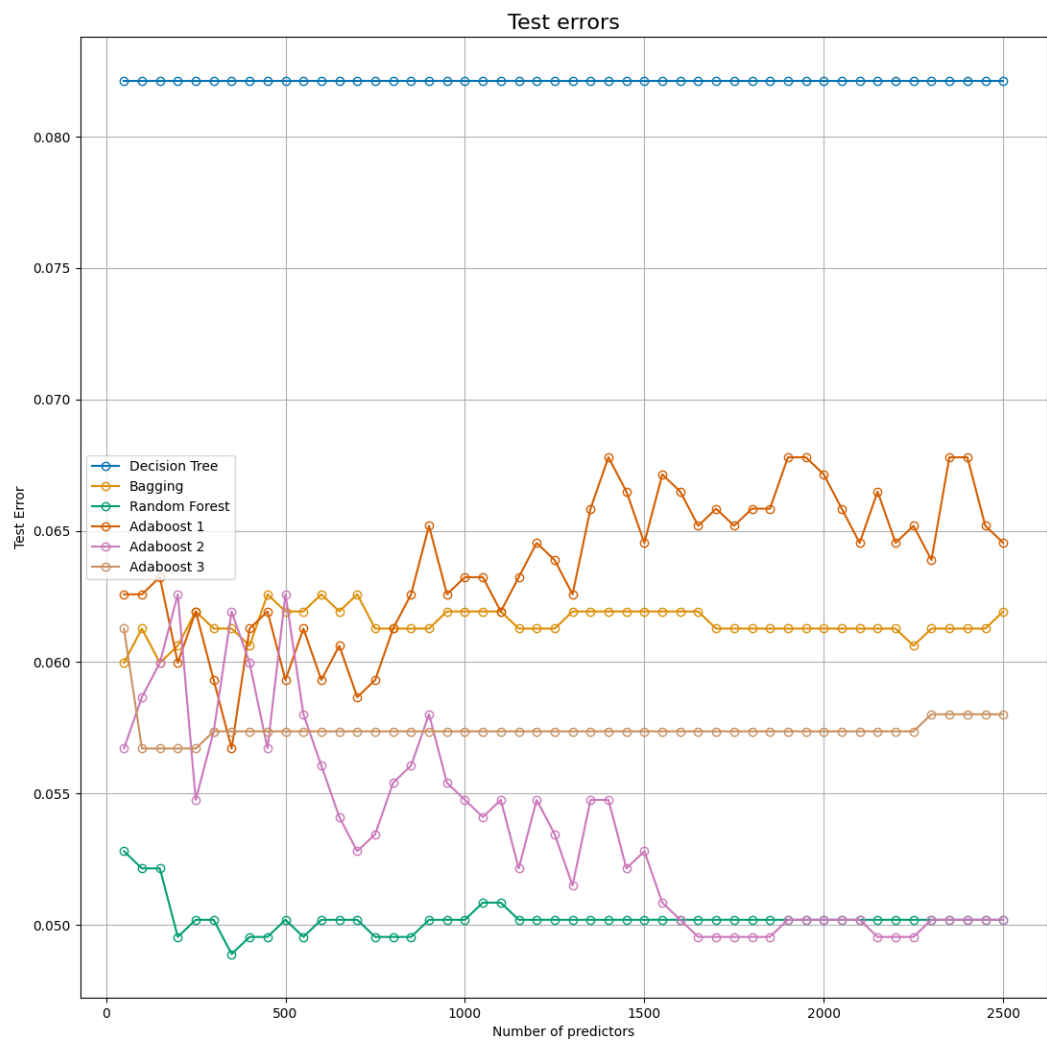
Test errors

The decision tree classifier with the lowest test error was at a max_leaf_nodes value of 96. The test error here was 0.0821382. The decision tree classifier cross-validation error generally drops until 96 max leaves, then starts increasing until it plateaus at 187 max leaves, and cv error of 0.0925958.

Decision Tree Classifier Cross-Validation Errors vs Max Number of Leaves